EDITORIAL



Editorial: Evidence on Questionable Research Practices: The Good, the Bad, and the Ugly

George C. Banks¹ · Steven G. Rogelberg¹ · Haley M. Woznyj¹ · Ronald S. Landis² · Deborah E. Rupp^{3,4}

Published online: 25 June 2016

© Springer Science+Business Media New York 2016

Abstract

Purpose Questionable research or reporting practices (QRPs) contribute to a growing concern regarding the credibility of research in the organizational sciences and related fields. Such practices include design, analytic, or reporting practices that may introduce biased evidence, which can have harmful implications for evidence-based practice, theory development, and perceptions of the rigor of science.

Design/Methodology/Approach To assess the extent to which QRPs are actually a concern, we conducted a systematic review to consider the evidence on QRPs. Using a triangulation approach (e.g., by reviewing data from observations, sensitivity analyses, and surveys), we identified the good, the bad, and the ugly.

Findings Of the 64 studies that fit our criteria, 6 appeared to find little to no evidence of engagement in QRPs and the other 58 found more severe evidence (91 %).

Implications Drawing upon the findings, we provide recommendations for future research related to publication practices and academic training.

Originality/value We report findings from studies that suggest that QRPs are not a problem, that QRPs are used at

a suboptimal rate, and that QRPs present a threat to the viability of organizational science research.

Keywords Questionable research practices QRPs · Research methodology · Philosophy of science · Ethics · Research methods

Introduction

Concerns exist regarding the credibility of research in the social and natural sciences (Cortina 2015; Kepes and McDaniel 2013; Nosek et al. 2015; Schmidt and Hunter 2015). These concerns are linked, in part, to the use of questionable research or reporting practices (QRPs). QRPs have been defined as "design, analytic, or reporting practices that have been questioned because of the potential for the practice to be employed with the purpose of presenting biased evidence in favor of an assertion" (Banks et al. 2016, p. 3). Examples of commonly discussed QRPs include selectively reporting hypotheses with a preference for those that are statistically significant, "cherry picking" fit indices in structural equation modeling (SEM), and presenting post hoc hypotheses as if they were developed a priori (Banks and O'Boyle 2013; John et al. 2012). Other typical QRPs might include reporting that a p value of 0.054 is p < 0.05 rather than p = 0.05, as well as adding and removing data and control variables in order to turn null results into statistically significant ones (Banks et al. 2016; John et al. 2012). These practices can occur with or without intent to deceive, but exist out of normative assumptions around research. By their presence in the literature, QRPs may harm the development of theory, evidence-based practice, and perceptions of the rigor and relevance of science. Herein, we review the available



[☐] George C. Banks gbanks3@uncc.edu

University of North Carolina at Charlotte, Belk College of Business, 9201 University City Blvd, Charlotte, NC 28223, USA

Illinois Institute of Technology, Chicago, USA

³ Purdue University, West Lafayette, Indiana, USA

⁴ University of the Western Cape, Bellville, Cape Town, South Africa

evidence from the social sciences in order to make conclusions about whether, given what we know to date, such concerns are warranted.

We review the evidence on methodological design and analysis of QRPs in a systematic fashion searching for evidence of the good, the bad, and the ugly. In other words, we looked for instances where QRPs seem not to be a problem (the good), where QRPs are used at a suboptimal rate, but perhaps are not overly problematic (the bad), and finally, we looked for evidence that QRPs represented a serious threat to the inferences made based on reported results (the ugly). We focus primarily on the organizational sciences and related social science fields such as education, political science, and accounting.

Following best practices for a systematic search (Kepes et al. 2013; Reed and Baxter 2009), we conducted a search in December 2015 using primarily Google Scholar and ProQuest Dissertations in order to identify both published and unpublished studies. We also searched for working papers at the National Bureau of Economic Research (http://www.nber.org/papers.html) and Social Science Research Network (http://www.ssrn.com/en/). First, databases were searched using combinations of the following keywords: (1) questionable research practices, (2) questionable reporting practices, (3) QRP, (4) HARKing, (5) phacking, (6) p-curve, (7) outcome-reporting bias, (8) underreporting, and (9) research ethics. Second, in addition to searching through the databases, we also conducted a citation search using references identified in Google Scholar. This involved backward- and forward-reference searches where we examined older studies cited by our identified studies and newer studies that cited our identified studies. Third, we submitted a call for published and unpublished studies over listservs, such as those sponsored by the Organizational Behavior, Human Resources, and Research Methods divisions of the Academy Management.

We limited our search to the social sciences due to the levels of criticism of late directed toward the social and organizational sciences (for reviews see Banks et al. 2016; Kepes and McDaniel 2013; Schmidt and Hunter 2015). Given the differences between the social and natural sciences regarding research methodologies (e.g., experimental designs, iterative research processes), there were concerns that the findings of one might not generalize to another. Furthermore, because of our interest in actual levels of engagement in methodological design-, analytic-, and reporting-QRPs, we excluded studies that focused (as a topic of study) on the treatment of human subjects, plagiarism, sample-level publication bias (i.e., entire samples are missing from the literature; see Kepes et al. 2012), simulations, replications, studies that only used hypothetical scenarios surrounding QRPs (e.g., vignettes) as opposed to considering actual behavior, or studies that focused on individual cases (e.g., retractions). In total, we identified 64 studies through this search. Despite our exhaustive search, we cannot rule out the possibility that a systematic difference may exist between studies that were available for identification compared to those that were not. Given the context, it may be that studies reporting a higher prevalence of engagement in QRPs were more likely to be identified by our search.

Review of Existing Evidence

Our review used a *triangulation* approach. Triangulation is characterized as "multiple reference points to locate an object's exact position" (Jick 1979, p. 602). All methodological approaches have limitations and are only as accurate as their underlying assumptions. Triangulation approaches in research use particular methods to compensate for weaknesses in other designs (e.g., Harrison et al. 2014; Rogelberg and Laber 2002). Thus, this approach draws upon multiple study designs, settings, and samples to consider engagement in QRPs (Kepes et al. 2012; Sackett and Larson 1990). Hence, our approach was holistic and allowed the consideration of many types of QRPs.

In the current review, we consider four primary types of evidence. First, we begin with a review of evidence from behavioral observations. This methodology primarily focuses on investigating how unpublished, raw studies in the form of protocols, dissertations, and conference papers transform into published journal articles. Second, we consider evidence from sensitivity analyses. These studies consider the probability of certain results and statistics appearing in journal articles. Third, we review evidence from self-report survey research where people indicate their own engagement in QRPs. Finally, we examine observer reports through survey research where people indicate the extent to which they have observed or know of others who have engaged in QRPs. Within each methodological category, we highlight examples of the research findings.

We summarize our findings in the Appendix, which provides the author, year, field, study topic, sample type, and key findings from each article that we reviewed. To the extent possible, we draw text regarding the key findings directly from the abstracts of each study with a focus on reporting results that highlight the extent to which QRPs are used. We encourage interested readers to refer to the primary studies for additional discussion of the nuanced results that are beyond the scope of our review. In text, we highlight studies that represent the range of findings identified as opposed to just focusing on those that are most exemplar.



Evidence from Behavioral Observations

A common technique used in behavioral observation studies of QRPs is to compare research protocols or early versions of a study (e.g., dissertations, conference papers) to the final paper that is published (O'Boyle et al. 2014; Pigott et al. 2013). The goal is to see if unsupported results were just as likely to appear in the final version as supported results. Further, one can also compare whether or not behaviors such as the removal of data, and adding/removing control variables was associated with turning a nonsignificant result into a significant one.

An advantage of the behavioral observation approach is that one does not have to be concerned with the potential for biased reporting due to social desirability as is the case when self- and observer-report surveys are used. A second advantage of this approach is that it is not dependent on the ability of researchers to recall engagement in QRPs that may have occurred years ago. A third advantage is that the technique is not concerned with researchers' perceptions of whether their behaviors are inappropriate or appropriate. Rather, the behavioral technique is focused on objectively describing how a paper changed over the course of its history.

That being said, this approach is not without limitations. For instance, there are many studies that are not available as protocols or unpublished manuscripts. Hence, the representativeness of samples used in this sort of research can be questioned. A second limitation of the behavioral approach is that one cannot determine if the motivation for engagement in QRPs was driven by authors or by reviewers and editors who may have pressured authors to use suboptimal research practices as a condition for publication, or a joint combination of the two. Further, it is not known if changes in the reported results may have been due to research practices improving as a result of editor/reviewer feedback and overall author development (which may occur in the event of student dissertations becoming junior faculty publications).

A total of 19 studies were identified that fit our criteria (see "Appendix" section). From these 19 studies, results suggest that although researchers engaged in QRPs to a varying extent, the influence of such practices appear to be severe. Of the 19 behavioral observation studies, 4 appeared to find little to no evidence of engagement in QRPs and the other 15 found more severe evidence. The most common forms of QRPs identified by the behavioral approach tend to be centered on an overabundance of significant findings (versus unsupported hypotheses) or on lax reporting practices with regard to methodological procedures, data cleaning, and/or data analysis. Here are a few examples to highlight the range of findings using this approach:

- When investigating the potential for data fabrication among undergraduates, Allen et al. (2015) found some evidence of inappropriate behavior. The authors concluded that there was a potential that the behavior was driven by a poor understanding of appropriate research methods and analysis.
- Bakker and Wicherts (2014) found no differences in median p-values when comparing studies that reported excluding outliers versus those that did not report dropping observations. Yet, this study did find that many studies do not report removing data despite the fact that reported statistics suggest that such removal did occur.
- O'Boyle et al. (2014) illustrated that when dissertations became published articles, the ratio of supported to unsupported hypotheses more than doubled (0.82:1 vs 1.94:1).
- After comparing conference papers and associated published journal articles, Banks et al. (2013) concluded that engagement in QRPs was infrequent relative to similar studies in the literature (e.g., O'Boyle et al. 2014; Mazzola and Deuling 2013; Pigott et al. 2013). However, when QRPs were used (e.g., data were removed; hypotheses that predicted a positive relationship were changed to a negative relationship), 34.5 % of unsupported hypotheses became supported, relative to just 13.2 % of supported hypotheses becoming unsupported.
- When looking across time, Fanelli (2012) found that, from 1990 to 2007, there was a 22 % increase in significant findings in research studies.

Evidence from Sensitivity Analyses

Sensitivity analyses can be used to evaluate engagement in QRPs by calculating the probability that a set of results is possible (Francis et al. 2014). As with the behavioral approach, sensitivity analyses have strengths and limitations. For instance, sensitivity analyses do not require researchers to answer truthfully on questionnaires, nor do researchers need to rely on respondents' memories of past behaviors. Sensitivity analyses are also not concerned with researchers' rationalizations of such behaviors, but rather focus on statistical probability estimations. Unlike the behavioral approach, one advantage of sensitivity analyses is that they do not require protocols or early drafts of a study in order to investigate engagement in QRPs. However, this approach can be limited. For instance, sensitivity analyses lose quite a bit of accuracy when attempting to establish the probability that a certain result was found in any individual study. Rather, sensitivity analyses are more



accurate when evaluating the probability of a set of results across hundreds of reported results.

A total of 14 studies were identified that fit our criteria and used sensitivity analyses (see "Appendix" section). Of these studies, none appeared to find little to no evidence of engagement in QRPs and the other 14 found more severe evidence. Considering the evidence from sensitivity analyses, it seems that p value manipulation is a widespread practice among the fields included in the current review. That is, a majority of the studies that employed sensitivity analyses suggest that researchers are incorrectly rounding p values or perhaps p-hacking to make their results seem "more significant" than they actually are. Below, we offer a few examples that highlight the range of findings:

- In their research on p values, de Winter and Dodou (2015) reported that dramatic increases of significant results may have been the result of both QRPs, but also improved methodological designs.
- After reviewing more than 30,000 articles, Hartgerink et al. (2016) reported direct evidence of p values being rounded incorrectly.
- Using a sample of over 250,000 p values reported in 20 years of research, Nuijten et al. (2015) found that
 - Half of all published psychology papers that use null hypothesis significance testing (NHST) contained at least one p value that was inconsistent with its test statistic and degrees of freedom.
 - One in eight papers contained a grossly inconsistent p value that may have affected the conclusion drawn.
 - The average prevalence of inconsistent *p* values has been stable over the years, or has declined.
 - The prevalence of gross inconsistencies was higher in *p* values reported as significant than in *p* values reported as nonsignificant.
- Leggett et al. (2013) found an overabundance of *p*-values immediately below the critical 0.05 threshold relative to other ranges, despite the finding that the probability of this result was unlikely. Further, the prevalence of this practice seems to have increased over the past 40 years. Several other studies reported similar results of unlikely levels of *p*-values immediately below the 0.05 threshold (Gerber and Malhotra 2008a, b; Masicampo and Lalande 2012).
- Despite low power, O'Boyle et al. (2015) found that most moderated multiple regression analyses identify statistically significant results. Further, while sample sizes have remained largely stable over time, the percent of significant results associated with tests of interactions has increased from approximately 42 % in the early 1990s to approximately 72 % in more recent research.

Evidence from Self-Report Surveys

The use of self-report surveys to investigate QRPs has several methodological strengths and limitations. First, given the degree of autonomy and discretion researchers have, there is a great deal of opportunity to engage in suboptimal research practices. In many cases, it is unlikely that even coauthors would be aware if inappropriate techniques were being used to manipulate results. Hence, self-report surveys are a way to identify engagement in QRPs that might not otherwise be observed. Surveys may also be used to investigate the extent to which engagement in QRPs is attributable to authors' own volitions compared to reviewer and editor requests in the review process. It may be the case that authors engage in inappropriate behavior in anticipation of mitigating reviewers' and editors' biases (Banks et al. 2016). Thus, surveys can help to sort out the motives behind engagement in ORPs and external pressures potentially associated with such practices. Relatedly, surveys can assist in disentangling how "questionable" some research behaviors really are for individual researchers. For instance, dropping an outlier, either for theoretical or methodological reasons, can change the conclusions one draws from the results. If a researcher has sound logic for this practice and is transparent, that practice is less questionable than if a researcher manipulates an analysis for the express purpose of turning a nonsignificant result into a statistically significant one. Carefully worded surveys can inform these sorts of issues.

Yet, there are also limitations to self-report surveys. The most obvious is that, even under conditions of confidentiality, researchers may not respond truthfully to survey questions due to socially desirable responding (Berry et al. 2012). Further, researchers may not be honest with themselves and may either answer that they did not engage in a practice or they might rationalize their behavior and make the argument that the behaviors were justified, even if they were not transparent in their reporting. Among those knowingly carrying out unethical practices, there is an incentive to under-report the use of QRPs so that such individuals might continue to keep such practices "under the radar." Thus, as with any method, there are advantages and disadvantages to the self-report survey when studying QRPs. One of the more problematic concerns may be the underreporting of QRP engagement (which is in many ways similar to the underreporting of counterproductive work behaviors in organizations; Berry et al. 2012). Thus, what we observe may be low-end estimates of QRPs.

A total of 17 studies were identified that fit our criteria and used self-report surveys (see "Appendix" section). Of these studies, 1 appeared to find little to no evidence of engagement in QRPs and the other 16 found more severe evidence. Many of the self-report studies tended to consider a range of



QRPs. Overall, though most studies employing self-report methods suggest that researchers are engaging in QRPs, the extent of engagement seemed to vary by QRP type. Taken as a whole, however, our review of the survey research indicates that QRPs are being used at a problematic rate. Here are a few examples that represent the range of findings:

- Bailey (2015) found a minimal association between researchers' acceptance of QRPs and the number of publications one has. For example, if researchers believe that the use of QRPs is appropriate, does that tend to mean they are more successful in publishing? While other studies have found a correlation between engagement in QRPs and publishing one's work in higher impact journals (Banks et al. 2013; O'Boyle et al. 2014), this study considered the issue more indirectly.
- John et al. (2012) reported that 45–50 % of the researchers surveyed stated that they engaged in selectively reporting results, 22–23 % reported having incorrectly reported *p* values, and 38–43 % reported having excluded data after considering how the practice would impact the results.
- Fiedler and Schwarz (2015) criticized past research and suggested that engagement in QRPs may be lower than has been implied and reported elsewhere. They argue that past research has asked if researchers ever engaged in a practice, while Fiedler and Schwarz focused more on how frequently researchers engage in such practices. They argue that their results suggest that base rates were lower than what has been found in other studies.
- Banks et al. (2016) found that about 11 % of researchers admitted to inappropriately reporting *p* values. Approximately 50 % of researchers said that they selectively reported results and presented post hoc findings as if they had been determined a priori. About a third of researchers surveyed reported engaging in post hoc exclusion of data and decisions to include/ exclude control variables to turn nonsignificant results into significant ones. The reporting of QRPs was not found to vary by academic rank.

Evidence from Observer-Report Surveys

Similar to the previously discussed methodological approaches, there are strengths and limitations to using observer-report surveys to study engagement in QRPs. Many QRPs may occur that cannot be identified via behavioral observations or sensitivity analyses. As with self-report surveys, one advantage of observer reports is that they can unearth those QRPs that can only be studied by asking researchers what occurred behind the scenes of data collection, analysis, and reporting of results. Another

advantage of using observer reports is that it reduces the potential for socially desirable responding (as compared to self-report surveys). Nonetheless, even observers in the form of coauthors or colleagues cannot observe and account for all analytic decisions made by other researchers. Thus, similar to self-reports, there is the potential for observer reports to provide underestimates of QRP frequency. While the observer-report approach is not perfect, it does provide complementary information to the other approaches described thus far.

A total of 14 studies were identified that fit our criteria and used observer-report surveys (see "Appendix" section). Of these studies, 1 appeared to find little to no evidence of engagement in QRPs and the other 13 found more severe evidence. Similar to the self-report surveys, the observer reports tended to investigate many QRPs within an individual study. Compared to the evidence from the self-report approach, observer reports paint an even grimmer picture of our scientific practices. The differences in results between the two survey approaches highlight the strengths and weaknesses of each method and illustrate the advantages of triangulation. Whereas people may be more reluctant to selfreport their own behaviors, they are willing to report when they have witnessed others engaging in QRPs. Results suggest that a large number of researchers are engaging in QRPs, though, like the self-report evidence, the extent of engagement varies by type. Here are a few examples that represent the range of findings uncovered:

- Bedeian et al. (2010) found that 79 % of researchers surveyed reported having observed others withholding methodological details or results. Ninety-two percent of respondents also reported having seen others present post hoc findings as those developed a priori and 78 % saw others selectively report findings.
- In another study focused on doctoral students, Banks et al. (2016) found that 12 % of doctoral student respondents indicated observing inappropriate reporting of p values, 55 % had seen selective reporting of results, and 58 % had seen the practice of reporting post hoc findings as a priori.
- In a meta-analysis of surveys asking about the behavior of colleagues, Fanelli (2009) found that 72 % of respondents reported observing a variety of QRPs, such as data manipulation.

Summary of the Good, the Bad, and the Ugly in QRP Research

We summarize our key findings in Table 1. In general, there were very few studies which identified little to no evidence for engagement in QRPs. It is not clear if this is because engagement in QRPs is ubiquitous, because of the



designs of the QRP studies, or because we had limited access to studies that found little to no engagement in QRPs.

The extent to which a finding is "bad" relative to "ugly" may depend on the practice itself as well as the frequency with which it is used. For instance, estimates of data falsification from self-reports are roughly 1-2 % (Banks et al. 2016; John et al. 2012). However, when observer reports are used, this number may be as large as 7 % (Banks et al. 2016), 14 % (Fanelli 2009), or even 27 % (Bedeian et al. 2010). Other levels of engagement in QRPs may be considered "bad," but less harmful, such as inappropriately rounding p values (Banks et al. 2016; John et al. 2012). Some QRPs, such as presenting a post hoc hypothesis as a priori, likely occur at more alarming rates (Banks et al. 2016; Bosco et al. 2015; John et al. 2012; Kerr and Harris 1998). Further, evidence of outcome-reporting bias seems to indicate that the practice is quite prevalent (John et al. 2012; Mazzola and Deuling 2013; O'Boyle et al. 2015; O'Boyle et al. 2014; Pigott et al. 2013) and that editors and reviewers play a role in the prevalence of this practice (Banks et al. 2016; LeBel et al. 2013). Additionally, though some studies found perhaps more mixed evidence of p-values occurring at high rates immediately below the traditional 0.05 threshold (Hartgerink et al. 2016; Nuijten et al. 2015), more found evidence that the practice was much more common (Leggett et al. 2013; Masicampo and Lalande 2012; Gerber and Malhotra 2008a, b).

When interpreting the good, the bad, and the ugly results from the current review, we want to note that there are many examples of sound research practice in our literature (e.g., Becker 2005; Locke 2007). Yet, engagement in QRPs is occurring at rates that far surpass what should be considered acceptable. Thus, some type of action is clearly needed to improve the state of our science. Below, we

provide some recommendations for improving publication practices and academic training.

Recommendations for Publication Practices and Academic Training

We believe the QRP discussion engenders a similar 'debate' as one sees in a discussion of climate change. For many years (nee, decades), scientists reported findings that indicated significant changes to the Earth's climate and were met with skepticism about whether the phenomenon was real, the degree to which climate change posed a significant problem, and whether human behavior was responsible. The current review is intended to provide a foundation upon which there can be agreement as to the extent that QRPs have been and are being practiced within the social and organizational sciences. Although the precise scope of the problem may be debated, there is sufficient evidence such that we cannot deny the significant presence of engagement in QRPs—the data do indeed triangulate. The challenges that remain before us are more about how we should best deal with QRPs.

While there are countless recommendations that can be made to address engagement in QRPs, we focus on those recommendations that we believe to be the most impactful. We do wish to note that we believe the challenge of QRPs is more of a "bad barrels" problem than a "bad apples" problem. That is, whereas there will always be individuals who violate responsible and ethical norms of research conduct, the majority of research to date suggests that our research systems inadvertently prime/reward the types of behaviors that derail our science (O'Boyle et al. 2014). Hence, our recommendations focus on addressing the issue of QRPs systematically. We summarize our recommendations in Table 2.

Table 1 Summary of key findings

Description

- 1 The vast majority of studies included in the current review identified evidence for the engagement in QRPs (91 % of studies)
- 2 Engagement in QRPs occurs at a nonideal rate and the current rates are likely underestimates
- 3 The extent to which QRPs are problematic varies by type and engagement frequency
- 4 Observer reports of QRPs typically illustrated higher prevalence rates than self-reports of engagement in QRPs
- 5 Nonsurvey-based methods, such as behavioral observations or sensitivity analyses, also provided consistent evidence of engagement in QRPs
- 6 Some of the more common QRPs include HARKing and selectively reporting results with a preference for those findings that are statistically significant
- 7 Editors and reviewers appear to play a role in the prevalence of QRPs
- 8 Engagement in QRPs has not been shown to vary by academic rank
- 9 The vast majority of QRP research has focused primarily on practices that affect p-values; more work is needed that investigates other types of QRPs



Changes to How We Review and Reward

First, we recommend that journals be more explicit about what sorts of research practices are and are not acceptable, and that they hold authors accountable for following journal policy. Despite the evidence that exists regarding engagement in QRPs, a recent review highlighted the fact that many journals in applied psychology and management, for instance, do not have explicitly stated policies pertaining to the vast majority of the QRPs reviewed in the current study (Banks et al. 2016). This could be easily rectified through the adoption of simple policy statements, and the requirement that submitting authors acknowledging (e.g., by checking boxes during the submission process) that they did not engage in multiple, separately, and explicitly described QRPs.

Second, we acknowledge that authors may engage in QRPs largely due to the pressures associated with publication. In particular, *p*-hacking, HARKing, selective reporting of results, and the like, are all encouraged by publication practices that implicitly reward the finding of 'significant results that confirm study hypotheses. Publication models such as Registered Reports or Hybrid Registered Reports address such practices by having authors only submit 'proposals' (e.g., https://cos.io/prereg/). That is, the review process is initially results blind. Manuscripts are evaluated on the basis of their theoretical and/or conceptual foundations and proposed methodologies. In the case of registered reports, *in-principle acceptances* may be offered to studies that are submitted prior to the submission of results and discussion sections.

The advantage of these types of submission models is that authors recognize that the quality of their research questions, hypotheses, and methodology will be evaluated independent of the research results. Thus, for example, if a researcher submitted a compelling and well-designed study as a (hybrid) registered report, which yielded null results, their chance of publishing the study should not be harmed. This approach should therefore serve to temper incentives for engaging in QRPs. Such submission models should also lead to more accurate/less biased reviewer ratings, given that reviewers have been shown to be more critical of research methodologies when null results are present (Emerson et al. 2010).

Several journals in management and applied psychology have begun to offer these sorts of review options for authors (for details see https://jbp.uncc.edu/). Nonprofit organizations, such as *The Center for Open Science* (https://cos.io/), have offered individual researchers the opportunity to preregister studies independent of journals and even offered 1000 research teams \$1000 for successfully publishing preregistered research in order to promote the initiative (https://cos.io/prereg/). In general, journals should also be more accepting of studies with nulls results. Perhaps more special issues on nulls results, such as the effort by the *Journal of Business and Psychology* are warranted (Landis et al. 2014).

As a third major approach to dealing with the engagement in QRPs, journals might also seek to increase the diversity of research that is published. Rather than an almost exclusive emphasis on papers that conform to the hypothetico-deductive model, editors and reviewers could

Table 2 Summary of key recommendations

Description

Changes to how we review and reward

- 1. Journals should be more explicit about what sorts of research practices are and are not acceptable, and hold authors accountable for following journal policy
- 2. Journals should consider the implementation and evaluation of the effectiveness of Registered Reports or Hybrid Registered Reports
- 3. Journals should be more accepting of null results
- 4. Journals should seek to increase diversity of research that is published to include more exploratory, inductive research as well as research based on abduction
- 5. Journals should be more welcoming of replication research

Changes to how we train students

- 6. Doctoral programs should emphasize cultures that promote open science and collaboration while discouraging engagement in QRPs
- 7. Research ethics training in research methods classes should expand beyond a sole focus on the protection of human subjects to also include discussions regarding QRPs

Conduct more research

8. There is still more we need to understand about QRPs. To date there has been a focus primarily on practices that affect *p*-values. More work is needed that examines other types of QRPs, such as in structural equation modeling, specification of priors in Bayesian statistics, or misreporting data in qualitative research



be more welcoming of papers built upon inductive reasoning (for a review see Spector et al. 2014). More specifically, some have lamented the potential overemphasis on theory for a priori findings without allowing the opportunity for interesting results to ultimately lead to the advancement of theory (Hambrick 2007). Locke (2007) stated that such policies among journals "encourages—in fact demands, premature theorizing and often leads to making up hypotheses after the fact—which is contrary to the intent of the hypothetico-deductive model" (p. 867). Exploratory, inductive research has led to the development of many well-known theories, such as goal-setting theory (Locke and Latham 2002) and social cognitive theory (Bandura 2001). Consequently, inductive research should be encouraged by journals as well as abductive approaches to research (Aliseda 2006). In general, journal editors could be more inclusive of different types of studies and correspondingly match their reviewer rating forms, examples, and exemplars—and furthermore, reviewers could be trained to welcome broader types of research.

In the end, well-conducted impactful research, in the many forms it can come, should be what we value (and publish). We have to make sure that our publication practices ensure that this is the case. We believe that (1) innovations to the review process, (2) promotion of inductive and abductive research, and (3) emphasis on publishing high-quality null results are three of the most critical steps that journal editors can take. The preceding points aside, there are many other tangible changes that can be made to our publication practices. For instance, principles such as those comprising the Editor Ethics code of conduct (https://editorethics.uncc.edu/) encourage implementing practices to reduce engagement in QRPs among action editors, reviewers, and authors. Further, journals may consider policies that promote open science and sharing among researchers by following the Transparency and Openness Guidelines (Nosek et al. 2015).

Changes to How We Train Students

To this point, our recommendations have largely focused on editorial policies. This emphasis is because editors and reviewers (including promotion and tenure reviewers) act as critical gatekeepers and so we believe that they have great responsibility to promote positive change (Rupp 2011). In other words, it is our general contention that authors will ultimately align with whatever editors and reviewers reward. That being said, we believe that authors still have important responsibilities to engage in sound, scientific practices, and codes of ethics exist to provide such guidance (see http://www.apa.org/ethics/code/). At the same time, scholars must constantly engage in self-

development exercises to ensure their personal competence in making the right decisions and being able to effectively evaluate others research. Programs such as the Center for the Advancement of Research Methods and Analysis (CARMA) could serve to improve students' (as well as their mentors' and instructors') understanding and use of statistics, such as *p*-values and fit indices.

Conduct More Research

Finally, there is still more we need to understand about QRPs. We note that most QRP research to date has focused primarily on practices that affect *p* values and that more work is needed that investigates other types of QRPs, such as, for example, fit indices in SEM (Banks and O'Boyle 2013), the specification of priors in Bayesian statistics (Banks et al. 2016), or misreporting interview results in qualitative research. Research has indicated that engagement in QRPs occurs when implementing null hypothesis significance testing (NHST), but it is not clear the extent to which engagement in QRPs is problematic for these other research approaches. We concur that research is sorely needed to evaluate the effectiveness of all the recommended strategies for reducing QRPs suggested herein.

Conclusion

The current study conducted a search of the literature on methodological design, analytic, and reporting of research practices that are questionable in nature. Of the 64 studies that fit our criteria, 6 appeared to find little to no evidence of engagement in QRPs and the other 58 found more severe evidence (91 %). Each of the studies reviewed had limitations associated with the various methods they employed. However, our triangulation approach allows us to have greater confidence that the findings uncovered are robust. Based on this analysis, we conclude that it is unlikely that most researchers engage in QRPs every time a study is conducted. For instance, if a team of researchers designs a study and finds support for most of their hypotheses, it is doubtful that there is motivation or a need to engage in QRPs. Yet, if initial support is largely not found, given the time, money, and energy that went into conducting a study and the enormous pressure from the current incentive system to publish, it is likely that researchers begin to consciously or subconsciously tinker with their analyses, their processes, and their reporting in order to present the best possible story to reviewers—to win the publishing "game." We hope that this review and our subsequent recommendations serve to advance a collegial dialogue on QRPs and to promote tangible and needed change.



Appendix

Author	Field	Topic	Sample type	Key findings
Evidence from be	ehavioral observati	ons		
Allen et al. (2015, p. 1)	Psychology	Data fabrication	Undergraduate students	Partial duplicates of data were identified; most possible explanations do not suggest nefarious intent
Bakker and Wicherts (2014, p. 1)	Psychology	Outlier removal	Journal articles	(1) Results showed no significant difference between the articles that reported excluding outliers and articles that did not in terms of median <i>p</i> -value, sample sizes, or prevalence of all reporting errors, large reporting errors, and reporting errors that concerned the statistical significance. (2) However, the study did find a discrepancy between the reported degrees of freedom of t-tests and the reported sample size in 41 % of articles that did not report removal of any data values. This suggests common failure to report data exclusions (or missingness) in articles
Banks et al. (2013, p. 1)	Management	Various QRPs	Conference papers	Engagement in QRPs were rare; yet, when such practices did occur, 34.5 % of unsupported hypotheses became supported compared to just 13.2 % of supported hypotheses becoming unsupported
Bosco et al. (2015, p .1)	Management/ psychology	HARKing	Effect sizes in journal articles	Correlations are significantly larger when hypothesized compared to nonhypothesized
Davis et al. (2007, p. 395)	Various	Causes of research misconduct	Case files of Office of Research Integrity	Causal factors implicated in research misconduct included: (1) personal and professional stressors, (2) organizational climate, (3) job insecurities, (4) rationalizations, (5) personal inhibitions, (6) rationalizations and, (7) personality factors
Fanelli (2012, p. 891)	Various	Outcome reporting bias	Journal articles	(1) The overall frequency of positive support has grown by over 22 % between 1990 and 2007, with significant differences between disciplines and countries. (2) The U.S. had published, over the years, significantly fewer positive results than Asian countries (and particularly Japan,) but more than European countries (and in particular the United Kingdom)
Field et al. (2016, p. 1)	Psychology	Various QRPs	Effect sizes	Manipulation of nonsignificant results to surpass a statistical threshold does not threaten meta-analytic inferences
Field et al. (2016, p. 1)	Psychology	Various QRPs	Effect sizes	Manipulation of nonsignificant results to surpass a statistical threshold may affect up to 19 % of research findings in I-O psychology
Field et al. (2015, p. 1)	Management/ psychology	Conflict of interest	Journal articles	(1) Effect size magnitude is not impacted by the presence or source of research funding across broad bivariate relation type. (2) Funded studies have a higher proportion of statistically significant findings (69 % of comparisons) and were also characterized by larger sample sizes (75 % of comparisons). (3) The pattern of results supports a methodological enhancement explanation for the funding–finding relation rather than a QRP-based explanation
Franco et al. (2016; p. 8)	Psychology	Outcome reporting bias	PROTOCOLS and journal articles	(1) 40 % of studies failed to fully report all experimental conditions and about 70 % of studies do not report all outcome variables included in the questionnaire. (2) Reported effect sizes are about twice as large as unreported effect sizes and are about 3 times more likely to be statistically significant
Masters (2012, p. iv)	Various	Fabrication, falsification	Cases of misconduct	The qualitative analysis indicated that 2.9 % of cases involved falsification, 4.4 % involved fabrication, and 4.4 % involved both fabrication and falsification
Matthes et al. (2015; p. 193)	Communication	Various QRPs	Journal articles	There were indications of small and insufficiently justified sample sizes, a lack of reported effect sizes, an indiscriminate removal of cases and items, an increasing inflation of p -values directly below $p < 0.05$, and a rising share of verified (as opposed to falsified) hypotheses



Author	Field	Topic	Sample type	Key findings
Mazzola and Deuling (2013, p. 279)	Psychology	Outcome reporting bias	Dissertations and journal articles	There was approximately 40 and 30 % differences between the two types of publications on their percentages of supported and unsupported hypotheses, respectively
Nagel (unpublished, p. 1)	Psychology	Post hoc exclusion of data	Journal articles	(1) Overall p -values within a sample of 70 articles on priming and automaticity, clustered around $p=0.05$. There was no systematic difference in statistical outcomes between studies that did and studies that did not exclude observations. (2) Reporting errors occurred in over half of all papers under investigation. (3) Exclusion of observations was not predictive of the number of reporting errors
O'Boyle et al. (2014; p. 1)	Management	Various QRP	Journal articles and dissertations	From dissertation to journal article, the ratio of supported to unsupported hypotheses more than doubled (0.82 to 1.00 versus 1.94 to 1.00).
Pigott et al. (2013, p. 1)	Education	Outcome reporting bias	Journal articles and dissertations	Nonsignificant outcomes were 30 % more likely to be omitted from a published study than statistically significant ones
Trainor (2015, p. x)	Education	Outcome reporting bias	Journal articles and dissertations	(1) Nonstatistically significant outcomes were 26 % more likely to get suppressed than statistically significant outcomes among individuals holding faculty positions and 50 % more likely among nonfaculty researchers. (2) When samples are predominantly white, nonstatistically significant outcomes are 24 % more likely to be suppressed when compared to 73 % among predominantly nonwhite samples. (3) Also, nonstatistically significant outcomes are 25 % more likely to be withheld among high school samples and 32 % more likely to be withheld among non-high school samples
Vasilev (2013, p. 717)	Psychology	Outcome reporting bias	Journal articles	The results indicated that almost all (95.4 %) articles considered found support for at least one tested hypothesis. 73 % of papers found support for all tested hypotheses
Veldkamp et al. (2014, p. 1)	Psychology	Reporting errors	Researchers	Overall, 63 % of the articles contained at least one <i>p</i> -value that was inconsistent with the reported test statistic and the accompanying degrees of freedom, and 20 % of the articles contained at least one <i>p</i> -value that was inconsistent to such a degree that it may have affected decisions about statistical significance. Overall, the probability that a given <i>p</i> -value was inconsistent was over 10 %
Evidence from sens	itivity analyses			
de Winter and Dodou (2015, p. 1)	Various	Distribution of <i>p</i> -values	Journal articles	(1) The p -values near the significance threshold of 0.05 on either side have both increased but with those p -values between 0.041 and 0.049 having increased to a greater extent (2013-to-1990 ratio of the percentage of papers = 10.3) than those between 0.051 and 0.059 (ratio = 3.6). (2) Contradictorily, $p < 0.05$ has increased more slowly than $p > 0.05$ (ratios = 1.4 and 4.8, respectively), while the use of "significant difference" has shown only a modest increase compared to "no significant difference" (ratios = 1.5 and 1.1, respectively). (3) Results are too inconsistent to draw conclusions on cross-cultural differences (e.g., U.S., Asia, and Europe). (4) The observed longitudinal trends are caused by negative factors, such as an increase of QRPs, but also by positive factors, such as an increase of quantitative research and structured reporting
Fanelli (2010, p. 1)	Various	HARKing	Journal articles	These results support the hypothesis that competitive academic environments increase not only scientists' productivity, but also their bias
Francis (2014, p. 1180)	Psychology	Success rates of studies	Journal articles	In total, problems with excess success rates appeared for 82 % (36 out of 44) of the articles in Psychological Science that had four or more experiments and could be analyzed



Author	Field	Topic	Sample type	Key findings
Francis et al. (2014, p. 1)	Psychology	Success rates of studies	Journal articles	The analyses indicated excess success for 83 % (15 out of 18) of the article in Science that report four or more studies and contain sufficient information for the analysis
Gerber and Malhotra (2008a, p. 3)	Political science	<i>p</i> -values	Journal articles	p-values were more common immediately below 0.05
Gerber and Malhotra (2008a, p. 3)	Sociology	<i>p</i> -values	Journal articles	p-values were more common immediately below 0.05
Hartgerink et al. (2016, p. 1)	Psychology	<i>p</i> -values	Journal articles	 (1) p-values were more common immediately below 0.05; the bump did no increase over the years and disappeared when using recalculated p-values (2) clear and direct evidence was found for the QRP "incorrect rounding o p-values"; (3) although one of the measures suggests the use of QRPs in psychology, it is difficult to draw general conclusions concerning QRPs based on modeling of p-value distributions
Head et al. (2015, p. 1)	Various	<i>p</i> -values	Journal articles	Manipulation of nonsignificant results to surpass a statistical threshold is widespread throughout science; results suggests that this manipulation probably does not drastically alter scientific consensuses drawn from meta analyses
Krawczyk (2015; p.1)	Psychology	p-values	Journal articles	(1) Some authors choose the mode of reporting in such a way that makes their findings seem more statistically significant than they really are; (2) they frequently report p -values "just above" significance thresholds directly, whereas other values are reported by means of inequalities (e.g. " $p < 0.1$ "), they round the p -values down more eagerly than up and appeat to choose between the significance thresholds and between one- and two sided tests only after seeing the data. (3) About 9.2 % of reported p -value are inconsistent with their underlying statistics (e.g., F or t) and it appear that there are "too many" "just significant" values
Leggett et al. (2013, p. 2303)	Psychology	<i>p</i> -values	Journal articles	(1) The frequency of <i>p</i> -values at and just below 0.05 was greater than expected compared to <i>p</i> -frequencies in other ranges. (2) While this overrepresentation was found for values published in both 1965 and 2005 it was much greater in 2005. (3) <i>p</i> -values close to but over 0.05 were more likely to be rounded down to, or incorrectly reported as, significant in 2005 than in 1965
Masicampo and Lalande (2012, p. 2271)	Psychology	<i>p</i> -values	Journal articles	p-values were more common immediately below 0.05
Nuijten et al. (2015, p. 1)	Psychology	<i>p</i> -values	Journal articles	(1) Half of all published psychology papers that use NHST contained at least one <i>p</i> -value that was inconsistent with its test statistic and degrees of freedom. (2) One in eight papers contained a grossly inconsistent <i>p</i> -value that may have affected the statistical conclusion. (3) The average prevalence of inconsistent <i>p</i> -values has been stable over the years or has declined. (4) The prevalence of gross inconsistencies was higher in <i>p</i> -values reported as significant than in <i>p</i> -values reported as nonsignificant
O'Boyle et al. (2015, p. 1)	Management	Outcome reporting bias	Journal articles	Despite low power, most MMR tests are statistically significant and while sample size has remained relatively stable over time, statistically significant MMR tests have risen from 42 % (1995–1999) to 52 % (2000–2004) to 60 % (2005–2009) to 72 % (2010–2014)
Schimmack (2014, p. 1)	Psychology	Various QRPs	Journal articles	The R-Index revealed the presence of QRPs when observed power is lowe than the rate of significant results



A		
AL	pendix.	continued

Author	Field	Topic	Sample type	Key findings
Self-reported surv	eys			
Bailey (2015, p. 1307)	Accounting	Various QRPs	Researchers	Only a small magnitude relation exists between acceptance of QRPs and publication count
Banks et al. (2016, p. 5)	Management	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Banks et al. (2016, p. 5)	Supply chain/sociology	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Bosco et al. (2015, p. 1)	Management/ psychology	HARKing	Researchers	Reported mixed reasons for occurrence of HARKing
Braun and Roussos (2012, p. 25)	Psychology	Various QRPs	Researchers	Degree of engagement in QRPs varies by type; North America was lower in almost all of the reported behaviors
Fanelli (2009, p. 1)	Various	Various QRPs	Researchers	(1) A pooled weighted average of 1.97 of scientists admitted to have fabricated, falsified, or modified data or results at least once—a serious form of misconduct by any standard—and up to 33.7 % admitted other QRPs. (2) Meta-regression showed that self-reports surveys, surveys using the words "falsification" or "fabrication," and mailed surveys yielded lower percentages of misconduct
Fiedler and Schwarz (2015, p. 1)	Psychology	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
John et al. (2012, p. 524)	Psychology	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Jørgensen et al. (2015, p. 1)	Software engineering	Outcome reporting bias	Conference attendees	Degree of engagement in QRPs varies by type
LeBel et al. (2013, p. 424)	Psychology	Methodological disclosure	Researchers	(1) Almost 50 % of contacted researchers disclosed the requested design specifications for the four methodological categories (excluded subjects, nonreported conditions and measures, and sample size determination). (2) Disclosed information provided by participating authors also revealed several instances of questionable editorial practices, which need to be thoroughly examined and redressed
List and Gallet (2001, p. 241)	Economics	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Martinson et al. (2009, p. 1491)	Various	Conflicts of interest	Researchers	(1) Being expected to obtain external funding and receiving federal research funding were both associated with significantly higher reports of 1 or more of 10 serious misbehaviors ($p < 0.05$) and neglectful or careless behaviors ($p < 0.001$). (2) Researchers with federal funding were more likely than were those without to report having carelessly or inappropriately reviewed papers or proposals (9.6 % vs. 3.9 %; $p < 0.001$). (3) Those with private industry involvement were more likely than were those without to report 1 or more of 10 serious misbehaviors (28.5 % vs. 21.5 %; $p = 0.005$) and to have engaged in misconduct (12.2 % vs. 7.1 %; $p = 0.004$); they also were less likely to have always reported financial conflicts (96.0 % vs. 98.6 %, $p < 0.001$)
Martinson et al. (2005, p. 737)	Various	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Martinson et al. (2006, p. 51)	Various	Various QRPs	Researchers	(1) When scientists believe they are being treated unfairly they are more likely to behave in ways that compromise the integrity of science. (2) Perceived violations of distributive and procedural justice were positively associated with self-reports of misbehavior among scientists



Author	Field	Topic	Sample type	Key findings
Necker (2014, p. 1747)	Economics	Various QRPs	Researchers	(1) Behavior such as data fabrication is (almost) unanimously rejected and admitted by less than 4 % of researchers. (2) Research practices that are often considered "questionable," e.g., strategic behavior while analyzing results or in the publication process, are rejected by at least 60 % of researchers. (3) Despite their low justifiability, these behaviors are widespread. (4) Ninety-four percent reported having engaged in at least one unaccepted research practice
Rajah- Kanagasabai and Roberts (2015, p. 1)	Various	Various QRPs	Undergraduates	Approximately one in seven students reported data fabrication and one in eight data falsification
Vul et al. (2009, p. 274)	Psychology	fMRI study accuracy	Researchers	(1) Past correlations are higher than should be expected given the reliability of both fMRI and personality measures. (2) Surveyed authors reported findings of this kind; more than half acknowledged using a strategy that computes separate correlations for individual voxels and reports means of only those voxels exceeding chosen thresholds. (3) Showed how this nonindependent analysis inflates correlations while yielding reassuring-looking scatter grams. (4) This analysis technique was used to obtain the vast majority of the implausibly high correlations in the survey sample
Observer report surve	eys			
Banks et al. (2016, p. 5)	Management	Various QRPs	Doctoral students	Degree of engagement in QRPs varies by type
Bedeian et al. (2010, p. 715)	Management	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Fanelli (2009, p. 1)	Various	Various QRPs	Researchers	In surveys asking about the behavior of colleagues, admission rates were 14.12 % for falsification, and up to 72 % for other QRPs
Glick and Shamoo (1994, p. 275)	Various	Various QRPs	Conference attendees	The vast majority of respondents had suspicions or evidence of other researchers performing questionable studies
Kattenbraker (2007, p. i)	Education	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Kerr and Harris (1998; p. 196)	Psychology/sociology	HARKing	Researchers	Two approaches to HARKing occurred at frequencies statistically indistinguishable from a more appropriate approach to hypothesis development
List and Gallet (2001, p. 241)	Economics	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Meyer and McMahon (2004, p. 413)	Accounting	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Riordan and Marlin (1987, p. 104)	Psychology	Data fabrication	Researchers	Participants perceived that data fabrication was relatively uncommon in the field
Swazey et al. (1993, p. 542)	Various	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Swazey et al. (1993, p. 542)	Various	Various QRPs	Doctoral students	Degree of engagement in QRPs varies by type



Author	Field	Topic	Sample type	Key findings
Tangney (1987, p. 62)	Various	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Titus et al. (2008, p. 980)	Various	Various QRPs	Researchers	Degree of engagement in QRPs varies by type
Wilson et al. (2007, p. 5)	Various	Research records and the resolutions of misconduct	Research integrity officers (RIO)	RIOs reported problems with research records in 38 % of the 553 investigations they conducted. Five types of poor record keeping practices accounted for 75 % of the problems with incomplete/inadequate records being the most common (30 %)

Key finding summaries were largely quoted from the abstracts (or text of the paper) of the articles and we provide the corresponding page numbers of the abstracts. Given space constraints in this table we report findings that were most applicable to the focus of the current review. We encourage readers to return to the primary studies for a more complete understanding of the research methodology as well as additional details on the findings of each study. HARKing = Hypothesizing after the results are known/presenting post hoc hypotheses as if they were developed a priori

References

- Aliseda, A. (2006). Abductive reasoning: Logical investigations into discovery and explanation. Dordrecht: Springer.
- Allen, P. J., Lourenco, A., & Roberts, L. D. (2015). Detecting duplication in students' research data: A method and illustration. *Ethics & Behavior*. doi:10.1080/10508422.2015.1019070.
- Bailey, C. D. (2015). Psychopathy, academic accountants' attitudes toward unethical research practices, and publication success. *The Accounting Review*, 90(4), 1307–1332.
- Bakker, M., & Wicherts, J. M. (2014). Outlier removal and the relation with reporting errors and quality of psychological research. PLoS One, 9(7), e103360.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. Annual Review of Psychology, 52, 1–26.
- Banks, G. C., et al. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management*, 42(1), 5–20.
- Banks, G. C., & O'Boyle, E. H. (2013). Why we need industrialorganizational psychology to fix industrial-organizational psychology. *Industrial and Organizational Psychology*, 6, 291–294.
- Banks, G. C., O'Boyle, E. H., White, C. D., & Batchelor, J. H. (2013). Tracking SMA papers to journal publication: An investigation into the phases of dissemination bias, Paper presented at the 2013 annual meeting of the Southern Management Association, New Orleans, LA.
- Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. Organizational Research Methods, 8, 274–289.
- Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education*, 9(4), 715–725. doi:10.5465/amle.2010.56659889.
- Berry, C. M., Carpenter, N. C., & Barratt, C. L. (2012). Do other reports of counterproductive work behavior provide an incremental contribution over self-reports? A meta-analytic comparison. *Journal of Applied Psychology*, 97, 613–636. doi:10.1037/ a0026739.
- Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2015). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology*. doi:10.1111/peps.12111.

- Braun, M., & Roussos, A. J. (2012). Psychotherapy researchers: Reported misbehaviors and opinions. *Journal of Empirical Research on Human Research Ethics*, 7(5), 25–29.
- Cortina, J. M. (2015). A revolution with a solution. Opening plenary presented at the meeting of the Society for Industrial/Organizational Psychology, Philadelphia, PA.
- Davis, M. S., Riske-Morris, M., & Diaz, S. R. (2007). Causal factors implicated in research misconduct: Evidence from ORI case files. Science and Engineering Ethics, 13(4), 395–414.
- de Winter, J. C. F., & Dodou, D. (2015). A surge of p-values between 0.041 and 0.049 in recent decades (but negative results are increasing rapidly too). *PeerJ*, 3, e733. doi:10.7717/peerj.733.
- Emerson, G. B., Warme, W. J., Wolf, F. M., Heckman, J. D., Brand, R. A., & Leopold, S. S. (2010). Testing for the presence of positive-outcome bias in peer review: A randomized controlled trial. *Archives of Internal Medicine*, 170, 1934–1939. doi:10. 1001/archinternmed.2010.406.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, 4(5), e5738.
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PloS One*, 5(4), e10271.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891–904. doi:10.1007/s11192-011-0494-7.
- Fiedler, K., & Schwarz, N. (2015). Questionable research practices revisited. Social Psychological and Personality Science, 7, 45–52.
- Field, J. G., Mihm, D., O'Boyle, E. H., Bosco, F. A., Uggerslev, K., & Steel, P. (2015). An examination of the funding-finding relation in the field of management. *Academy of Management Proceedings*. Paper presented at the Academy of Management Annual Meeting, Vancouver, Canada (p. 17463).
- Field et al. (2016). *The extent of p-hacking in I/O psychology*. Paper presented at the Society of Industrial/Organizational Psychology Annual Conference in Anaheim, CA.
- Francis, G. (2014). The frequency of excess success for articles in Psychological Science. *Psychonomic Bulletin & Review, 21*(5), 1180–1187.
- Francis, G., Tanzman, J., & Matthews, W. J. (2014). Excess success for psychology articles in the journal Science. *PLoS One*, 9(12), e114255.



- Franco, A., Malhotra, N., & Simonovits, G. (2016). Underreporting in psychology experiments evidence from a study registry. *Social Psychological and Personality Science*, 7(1), 8–12.
- Gerber, A., & Malhotra, N. (2008a). Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science*, 3, 313–326. doi:10.1561/100.00008024.
- Gerber, A. S., & Malhotra, N. (2008b). Publication bias in empirical sociological research do arbitrary significance levels distort published results? *Sociological Methods & Research*, 37, 3–30. doi:10.1177/0049124108318973.
- Glick, J. L., & Shamoo, A. E. (1994). Results of a survey on research practices, completed by attendees at the third conference on research policies and quality assurance. Accountability in Research, 3(4), 275–280.
- Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, 50, 1346–1352.
- Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle Jr., E. H., & Short, J. C. (2014). Publication bias in strategic management research. *Journal of Management*. doi:10.1177/0149206314535438.
- Hartgerink, C. H., van Aert, R. C., Nuijten, M. B., Wicherts, J. M., & van Assen, M. A. (2016). Distributions of p-values smaller than.05 in Psychology: What is going on? *PeerJ*, 4, e1935.
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, 13(3), e1002106.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. Administrative Science Quarterly, 24, 602–611. doi:10.2307/2392366.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. doi:10. 1177/0956797611430953.
- Jørgensen, M., Dybå, T., Liestøl, K., & Sjøberg, D. I. (2015). Incorrect results in software engineering experiments: How to improve research practices. *Journal of Systems and Software*,. doi:10.1016/j.jss.2015.03.065.
- Kattenbraker, M. (2007). Health education research and publication: ethical considerations and the response of health educators (Unpublished thesis). Southern Illinois University Carbondale, Carbondale, IL.
- Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, 15, 624–662. doi:10.1177/109442811245 2760.
- Kepes, S., & McDaniel, M. A. (2013). How trustworthy is the scientific literature in I-O psychology? *Industrial and Organi*zational Psychology: Perspectives on Science and Practice, 6, 252–268.
- Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic reviews in the organizational sciences: Two meta-analytic schools on the way to MARS (the Meta-Analytic Reporting Standards). *Journal of Business and Psychology*, 28, 123–143.
- Kerr, N. L., & Harris, S. E. (1998). HARKing: hypothesizing after the results are known: Views from three disciplines. Unpublished manuscript, Michigan State University, East Lansing.
- Krawczyk, M. (2015). The search for significance: A few peculiarities in the distribution of p-values in experimental psychology literature. *PloS One*, 10(6), e0127872.
- Landis, R. S., Lance, C. E., Pierce, C. A., & Rogelberg, S. G. (2014). When is nothing something? Editorial for the null results special issue of Journal of Business and Psychology. *Journal of Business and Psychology*, 29, 163–167. doi:10.1007/s10869-014-9347-8.

- LeBel, E. P., Borsboom, D., Giner-Sorolla, R., Hasselman, F., Peters, K. R., Ratliff, K. A., & Smith, C. T. (2013). PsychDisclosure.org grassroots support for reforming reporting standards in psychology. *Perspectives on Psychological Science*, 8(4), 424–432.
- Leggett, N. C., Thomas, N. A., Loetscher, T., & Nicholls, M. E. (2013). The life of p: "Just significant" results are on the rise. The Quarterly Journal of Experimental Psychology, 66(12), 2303–2309.
- List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, 20(3), 241–254
- Locke, E. A. (2007). The case for inductive theory building. *Journal of Management*, 33, 867–890.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*, 705–717.
- Martinson, B. C., Anderson, M. S., Crain, A. L., & De Vries, R. (2006). Scientists' perceptions of organizational justice and self-reported misbehaviors. *Journal of Empirical Research on Human Research Ethics*, 1(1), 51–66.
- Martinson, B. C., Anderson, M. S., & De Vries, R. (2005). Scientists behaving badly. *Nature*, 435(7043), 737–738.
- Martinson, B. C., Crain, A. L., Anderson, M. S., & De Vries, R. (2009). Institutions' expectations for researchers' self-funding, federal grant holding and private industry involvement: Manifold drivers of self-interest and researcher behavior. Academic Medicine: Journal of the Association of American Medical Colleges, 84(11), 1491–1499.
- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p-values just below. 05. The Quarterly Journal of Experimental Psychology and Aging, 65(11), 2271–2279. doi:10.1080/ 17470218.2012.711335.
- Masters, E. A. (2012). Research misconduct in National Science Foundation funded research a mixed-methods analysis of 2007-2011 research awards (Unpublished doctoral dissertation). Northcentral University, Prescott Valley, AZ.
- Matthes, J., Marquart, F., Naderer, B., Arendt, F., Schmuck, D., & Adam, K. (2015). Questionable research practices in experimental communication research: A systematic analysis from 1980 to 2013. Communication Methods and Measures, 9(4), 193–207.
- Mazzola, J. J., & Deuling, J. K. (2013). Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I-O journal articles. *Industrial and Organizational Psychology:* Perspectives on Science and Practice, 6(03), 279–284.
- Meyer, M. J., & McMahon, D. (2004). An examination of ethical research conduct by experienced and novice accounting academics. Issues in Accounting Education, 19(4), 413–442.
- Nagel, M., Wicherts, J. M., & Bakker, M. Participant exclusion in psychological research: A study of its effects on research results. *Unpublished manuscript*.
- Necker, S. (2014). Scientific misbehavior in economics. *Research Policy*, 43(10), 1747–1759.
- Nosek, B. A., et al. (2015). Promoting an open research culture: Author guidelines for journals to promote transparency, openness, and reproducibility. *Science*, 348, 1422–1425. doi:10.1126/science.aab2374.
- Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*. doi:10.3758/s13428-015-0664-2.
- O'Boyle, E. H., Banks, G. C., & Gonzalez-Mule, E. (2014). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management*,. doi:10.1177/014920 6314527133.



- O'Boyle, E. H., Banks, G. C., Carter, K., Walter, S., & Yuan, Z. (2015). A 20-year review of outcome reporting bias in moderated multiple regression. Paper presented at the annual meeting of the Academy of Management, Vancouver, British Columbia.
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*. doi:10.3102/0013189X13507104.
- Rajah-Kanagasabai, C. J., & Roberts, L. D. (2015). Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior. Frontiers in Psychology, 6, 1–11.
- Reed, J. G., & Baxter, P. M. (2009). Using reference databases. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 74–101). New York: Russell Sage Foundation.
- Riordan, C. A., & Marlin, N. A. (1987). Some good news about some bad practices. *American Psychologist*, 42(1), 104–106.
- Rogelberg, S. G., & Laber, M. (2002). Securing our collective future: Challenges facing those designing and doing research in Industrial and Organizational Psychology. In S. G. Rogelberg (Ed.), Handbook of research methods in industrial and organizational psychology (pp. 479–485). London: Blackwell.
- Rupp, D. E. (2011). Research and publishing ethics: Editor and reviewer responsibilities. *Management and Organizational Review*, 7, 481–493.
- Sackett, P. R., & Larson, J. R. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 419–489). Palo Alto, CA: Consulting Psychologists Press.
- Schimmack, U. (2014). Quantifying statistical research integrity: The Replicabilty-Index. Unpublished manuscript.

- Schmidt, F. L., & Hunter, J. E. (2015). Methods of meta-analysis: Correcting error and bias in research findings (3rd ed.). Newbury Park, CA: Sage.
- Spector, P. E., Rogelberg, S. G., Ryan, A. M., Schmitt, N., & Zedeck, S. (2014). Moving the pendulum back to the middle: Reflections on and introduction to the inductive research special issue of Journal of Business and Psychology. *Journal of Business and Psychology*, 29, 499–502. doi:10.1007/s10869-014-9372-7.
- Swazey, J. P., Anderson, M. S., Lewis, K. S., & Louis, K. S. (1993).
 Ethical problems in academic research. *American Scientist*, 81(6), 542–553.
- Tangney, J. P. (1987). Fraud will out-or will it? New Scientist, 115, 62–63.
- Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, 453(7198), 980–982.
- Trainor, B. P. (2015). Incomplete reporting: Addressing the problem of outcome-reporting bias in educational research (Unpublished doctoral dissertation). Loyala University, Chigao, IL.
- Vasilev, M. R. (2013). Negative results in European psychology journals. Europe's Journal of Psychology, 9(4), 717–730.
- Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PloS One*, 9(12), e114876.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3), 274–290.
- Wilson, K., Schreier, A., Griffin, A., & Resnik, D. (2007). Research records and the resolution of misconduct allegations at research universities. Accountability in Research, 14(1), 57–71.

